Issues Relating to the Sampling Design of the National Children's Study

May 2004

Issues Relating to the Sampling Design of the National Children's Study

Table of Contents
May 2004

Note: CD-ROMs are included in this mailing. The CD contains electronic copies of some of these same documents, in addition to more background information that you may find interesting or useful when preparing for the June 28–29 NCSAC meeting.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health
Centers for Disease Control and Prevention
U.S. ENVIRONMENTAL PROTECTION AGENCY

THE NATIONAL CHILDREN'S STUDY
HEALTH GROWTH ENVIRONMENT

MEMORANDUM

Date: May 17, 2004

To: National Children's Study Advisory Committee (NCSAC)

From: Donald R. Mattison, Chair

Subject: Tasks and Supporting Materials for the June Advisory Committee Meeting

As discussed at the last NCSAC meeting, a large portion of the meeting in June (Monday, June 28 and Tuesday, June 29 in Alexandria, VA) will be devoted to discussions and recommendations on sampling designs. Because of the complexity of this issue, we will be providing specific questions to the NCSAC, the answers to which will serve as advice for the next stages of the Study development. Also, we will provide you with background materials to assist in your deliberations.

In the interest of time, and to assist your thinking on these issues, we are forwarding materials to you in two stages. The first stage, including general issues and supporting material, is contained in this memo and attachments. The next stage of specific questions and presentations to be made at the NCSAC meeting are dependent upon several other activities in progress that are managed by the Program Office. For example, they are obtaining input on several specific issues, including the numbers of households/women needed, the degree of clustering needed to obtain community measures, the identification of alternative sampling frames, the costs for alternative approaches, and the potential effects of mobility. This information is needed to develop questions for the second stage. Specific questions for the NCSAC will be developed by an ad-hoc subcommittee of the NCSAC, Program Office, and ICC and provided for your consideration shortly before the June meeting.

For the first stage, we are sending you the following materials:

- Battelle Report Background Materials
  - Table of Contents for the entire report
  - Glossary of terms
  - Executive Summary for the Battelle Report
  - Chapter 1 of the Battelle Report
  - Appendix A to the Battelle Report is a white paper on the Advantages and Limitations of Alternative Sampling Methods for the Study
- Other items included in the Battelle Report are available on a CD included in this mailing. Please note that these are provided for your information only. The report and appendices are quite long and were developed for the Sampling Workshop, and are NOT essential to your preparation for the June NCSAC meeting.

- Report from the Sampling Workshop expert panel prepared by David Savitz with assistance from workshop panel. This report identifies two design alternatives, and discusses advantages and disadvantages for each.
- Bob Michael's exercise designed to elicit our preferences concerning Study goals and design.
- One of the slides presented by Rod Little at the workshop – this is included because it clearly describes some of the issues we will need to discuss.
- Summaries of some of the trade-offs to be considered for different design options or features.

Again, the major focus for the upcoming meeting is to discuss and recommend optimum sampling design options, or features of these designs (e.g., sampling frames, listing and selection methods, organizational structures, feasibility, cost, quality of data available for defining the relationship between "environment" and health and development) that should be considered by the Study planners as they design the National Children's Study. We request that you read the enclosed material and perform Bob Michael's exercise prior to the meeting. As you read the attachments, please keep in mind that we will be asked to make recommendations about the sampling design, including consideration of such issues as the:

- Identification and selection of geographic areas,
- Identification and selection of individuals within these areas, and
- Timing of enrollment.

Please consider trade-offs related to scientific merit, costs, and feasibility involved in alternative designs, especially those identified by the Sampling Workshop Panel.

Thank you for your participation in the upcoming meeting. Selection of the sampling design is a critical issue for the Study planners and a complicated one, given the broad range of disciplines involved and the differences in their perspectives, acceptable practices, and requirements. The NCSAC can serve a unique and valuable function in providing advice that leads to a Study sampling design that best meets its critical goals.

Enclosures:
Executive Summary Sampling Design
Battelle White Paper (Table of contents, Glossary, Chapter 1)
Battelle White Paper Appendix A
Sampling Panel Report
Bob Michael Exercise
Rod Little Design Issues

Executive Summary for the White Paper on Evaluation of Sampling Design Options for the National Children's Study[1]

OVERVIEW

The sampling design for recruiting women in either early stages of pregnancy or prior to conception into the National Children's Study (NCS) is one of the most difficult challenges facing NICHD and its Federal partners, CDC, EPA and NIEHS. With many competing objectives and multiple scientific hypotheses, the sampling design for the NCS defies being reduced into a one-dimensional optimization problem that is common to most other public-health research studies. The report does not attempt to develop a single optimal sampling strategy for the NCS. Rather, it establishes a conceptual framework for combining multiple modes of recruiting women into the study, and then compares and contrasts the performance of a range of design options under this framework with respect to retention of study subjects, cost of study implementation, and power to address the NCS core hypotheses. Thus, the report is intended to be used by study planners as a resource to help make informed choices on the sampling design for the NCS.

Prior to the development of an appropriate sampling approach, it is important to first consider the goals and statutory requirements of the study and the population of interest for the study. Broadly speaking, the main objective of the NCS is to study relationships between exposures, including chemical, physical, biological, and psychosocial exposures, and outcomes. As such, the aims of the NCS core hypotheses are to evaluate whether these exposures are associated with the occurrence of a disease, or changes in the associated outcome measures, so that appropriate actions (e.g., education on risk factors, or early detection of diseases) can be taken for the affected populations. Since the NCS will necessarily study contemporary children (children born in the United States during the NCS recruitment period), by the time conclusions are drawn from the NCS data, it will in most cases be too late to take effective action for this contemporary population. Thus, in the terminology of Deming (1953) and Hahn and Meeker (1993), we consider the NCS to be primarily an "analytical" study rather than an "enumerative" (or "descriptive") study.

Assuming first that the NCS will focus on a sample of contemporary children, we adopted the notion of an ideal target population that represents all children born in the U.S. during a specified recruitment period for the study. This allowed for the consideration of multiple sampling approaches, and evaluation of how well they cover the ideal target population. With the goal of recruiting women in early pregnancy and/or women of childbearing age prior to conception, we focused on three primary sampling models, a Household model, a Physician's Office model, and an Academic Medical Centers (broadly defined to include coordinating centers, medical centers, etc.) model, each having apparent advantages and disadvantages in light of the objectives of the NCS (e.g., coverage of the target population, screening requirements,

[1] This is a summary of the report provided to the sampling design workshop committee and does not address comments received at the workshop as a revised version of the report has not been developed.

ability to sample prior to pregnancy, ability to foster community involvement, ability to capitalize on pre-existing relationships with patients, etc.) and each having support from different members of the scientific community that have been involved in the planning process for the NCS.

This led to the consideration of dual- or multi-frame sampling strategies that would combine a broad probability-based population-wide sample, such as a national household sample or a household sample restricted to geographic regions which could be covered by qualified Centers, with a sample selected from patient lists of qualified Centers or physician's offices. By incorporating a sampling strategy based on the Household frame, the NCS may have a greater chance of being truly representative of the entire United States or of the selected areas. In addition, such a sample could ensure appropriate representation of low-income subjects or subjects from minority ethnicities using standard techniques for oversampling. However, the downside to this is that some of the subjects might be more likely to refuse to participate in the study, or might be more difficult to retain (i.e., be more likely to drop out before study completion). A careful choice of a more convenient frame, such as Center patients, can potentially identify a more compliant population (lower refusal rates, higher retention rates, easier tracking, greater cooperation with follow-up appointments, etc.). For example, study subjects recruited through an academic medical center already have built-in alternative tracking and contact mechanisms, as well as incentives to maintain contact with study staff as part of receiving ongoing care for their child. However, this frame is also likely to exclude certain segments of the population from the sample, such as women without access to healthcare.

In other words, a possible multi-frame sampling strategy for the NCS would combine all three models (Household, Physician's Office, and Centers) into an integrated framework. This use of multi-frame sampling is appealing from a heuristic perspective in terms of enhancing study validity by overcoming weaknesses associated with each approach (e.g., weaknesses in coverage, anticipated retention rates, efficiency, varying degrees of willingness to undergo burden, etc.); however, it does present a number of challenges associated with how data from the separate cohorts should be combined. For example, statistical analysis of data collected in such a manner poses considerable challenges, such as determining an appropriate approach to assigning sample weights to all study participants, as does determining the appropriate "mix" of the multiple frames given the numerous, and at times competing, objectives of the study (e.g., national probability-based sampling may provide greatest generalizability of the results but may result in relatively low retention rates over the course of the study).

In Chapter 3 of the report we describe an example multi-frame sampling approach by introducing a "Family of Designs" that combines the Household and Center-based models. Conceptually, this Family of Designs provides a multiple-approach solution for planning the study, in which part of the study population will be recruited in a manner that maximizes the opportunity for detailed and rigorous data collection, while another part of the study preserves the ability to generalize important study results to the population of interest. The intent is to maximize the advantages of different approaches while minimizing their limitations, resulting in a study design that is more optimal overall than one that is limited to a single recruitment approach. Additionally, the flexibility of a multi-frame approach may allow the study to more easily adapt if some approaches result in lower than anticipated (or unacceptable) response rates,

and may allow the study to satisfy many of the competing objectives of the NCS – without completely sacrificing any single objective for another.

Figure 1 displays a conceptual layout of the Family of Designs considered in the report. These designs initiate with identifying a fraction (P_1) of the NCS cohort that is recruited through a national probability-based sampling (NPBS) approach. Once this fraction is determined, the NPBS portion of the cohort is selected in a multi-stage clustered design where counties are the primary sampling units (PSUs), and households are sampled within counties (or other geographic units) to identify women of child-bearing age. (Note that other sampling frames, such as a physicians office frame, could also be considered for recruiting study participants within selected PSUs.) The remaining fraction of study subjects (1- P_1) are located within geographic regions corresponding to a set of purposively selected Academic Medical Centers and are recruited through a variety of mechanisms. Among the participants located within the Academic Medical Centers, we assume that a fraction (P_2) are recruited from a probability-based sample from areas in proximity to the Centers [e.g., from the metropolitan statistical areas (MSAs) surrounding the Centers], another fraction (P_3) are recruited from a probability-based sample of Center patients, and the remaining fraction (1- (P_2+P_3)) are recruited from an opportunity or convenience sample (see Chapter 3 for further details).

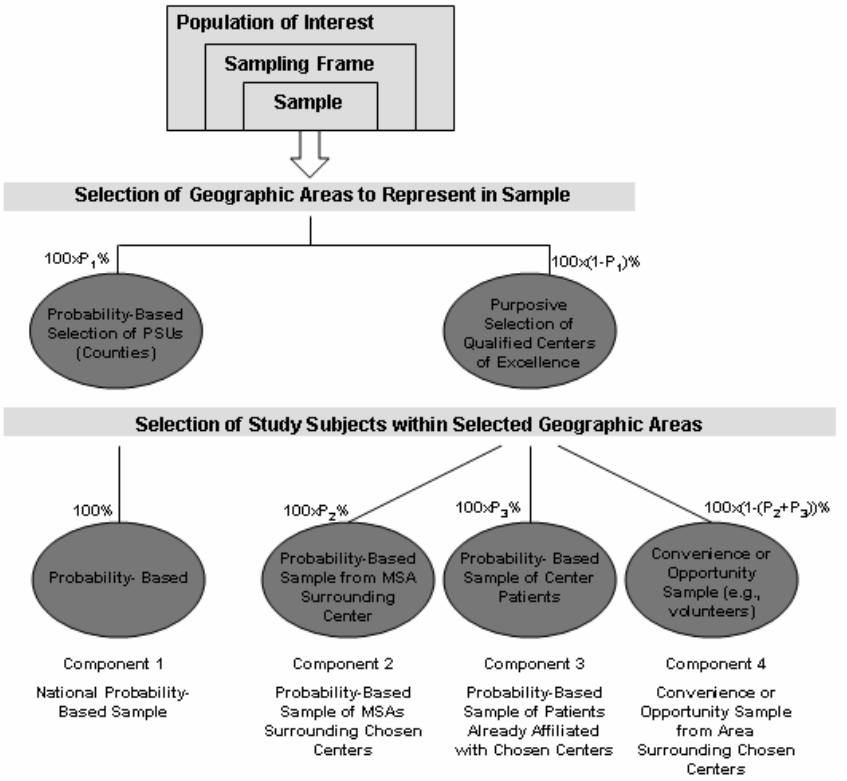

Figure 1. Conceptual Model for the Family of Designs.

Within this family of designs, there remain a large number of design possibilities. For example, what fraction of the cohort will be selected in the NPBS, how many PSUs will be

utilized, and what fraction of the Centers cohort should be selected using probability-based sampling of the area in proximity to the Centers? By specifying answers to these questions, candidate designs can be identified for more careful study of their corresponding characteristics. In order to focus on specific design examples when evaluating costs, statistical power, retention rates, etc., we consider a set of 23 designs (see Chapter 3) in which we allow the parameters involved in the Family of Designs (P_1, P_2, P_3, number of PSUs, etc.) to span a broad range of possible values so that an indication of the effect of changing these design parameters can be obtained and a more informed choice of design can be made. Chapter 5 of the report outlines the steps necessary in conducting the NPBS and Centers sampling approaches, and in combining the subjects sampled using these alternative approaches. As an example sample realization, Figure 2 displays a geographic representation for a realization of a design with 100 PSUs in the NPBS and 38 purposively selected Centers. The figure displays the counties selected in the NPBS (green counties), the counties that correspond to the MSA of one of the 38 purposively selected Centers (red counties), and the counties that were selected in the NPBS and correspond to the MSA of one of the purposively selected Centers (blue counties).

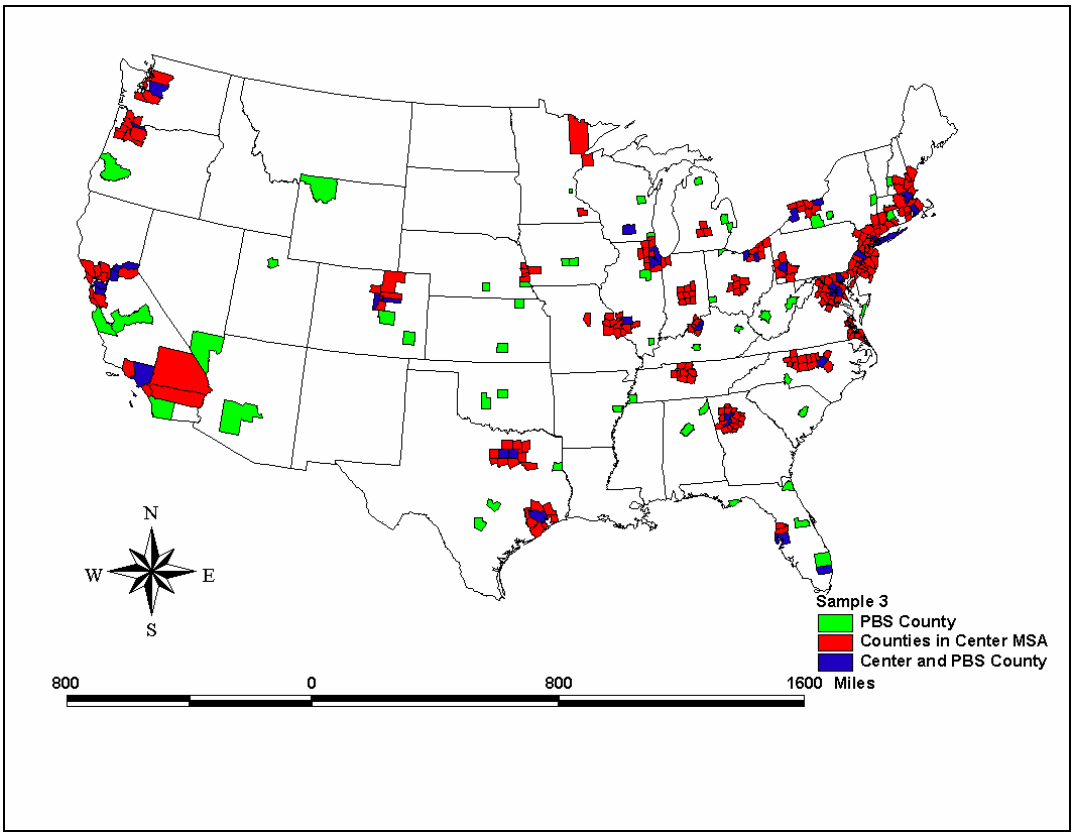

Figure 2. Geographic Representation for an Example Realization of a Design with 100 PSUs in the NPBS and 38 Purposively Selected Centers.

In order to estimate costs and conduct power analyses for important hypotheses another necessary design characteristic is the retention rate (i.e., the percentage of the original cohort that continues to participate in the study over time) associated with a given design. Retention rates have an effect on cost estimates since the number of children remaining in the study highly influences the costs of data collection. For power calculations, retention rates are also important,

especially when evaluating hypotheses that can be tested only after health effects are assessed in later stages of life. Chapter 7 and Appendix G of the report describe retention rates seen in other longitudinal studies, estimate retention rates based on these other studies, and outline the retention rate assumptions that are utilized in the cost estimates and power analyses.

As discussed in Chapter 7, it is important to note that no other studies involve the same scope, size, and complexity as that envisioned for the NCS, and, thus, estimating recruitment and retention rates based on these studies is very uncertain. Adding to this uncertainty is the effect of subject burden on retention rates and the difficulty in characterizing this burden given that the specific measurements and final NCS protocol have yet to be fully developed. Admittedly, it may be the case that recruitment and retention rates for the NCS will generally be higher than those observed in other studies (e.g., due to incentive programs, the important nature of the NCS, etc.), or it may be the case that recruitment and retention rates for the NCS will be lower than those observed in the other studies (e.g., due to subject burden, the length of the study, the methods of recruitment, etc.). To indicate the effect of the assumed retention rates on study costs and power to address research objectives, two approaches to estimating retention rates were presented. Based on data observed in other relevant studies[2], both approaches assumed that there would be differences in retention rates between study subjects that are recruited using probability-based sampling from relatively unrestricted populations compared to study subjects recruited using probability-based sampling from a much more restricted and convenient sampling frame or through convenience sampling. The first approach assumed a simple exponential decay model for retention rates experienced under different methods of recruiting study subjects into the NCS based on what was observed in historical studies. The second approach assumed an exponential decay model with the rate of decay experienced under the different methods of recruiting study subjects converging to a common value as the time of participation in the study increases. Figure 3 displays the assumed retention rates under these two different approaches with the left panel of the figure displaying the rates for a simple exponential decay retention model and the right panel displaying the rates for a retention model with a converging rate of decay. Also included on the graph for reference are the retention rates identified for other probability-based studies (denoted by a "P") and the retention rates identified for other Hospital/Center-based studies (denoted by a "*").

[2] Factors considered when selecting relevant studies included whether the study focused on young children, whether it focused on relevant health outcomes, whether it involved longitudinal follow-up, and/or whether it involved collection of biological/environmental samples or clinical/medical measures.

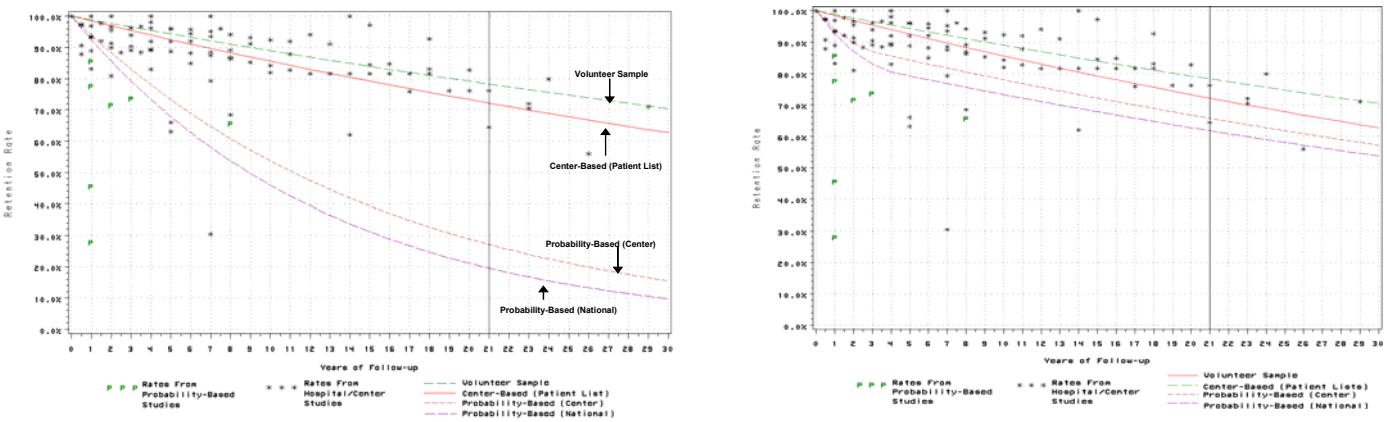

Figure 3. Retention Rates Observed in Other Similar Studies, and Assumed Retention Rates Under Two Different Estimation Approaches.

Using the assumptions regarding retention rates, cost estimates and power calculations were investigated for each of the selected 23 designs in order to evaluate important differences for these criteria within the family of designs. In characterizing cost and power estimates across the different designs, we considered two different design constraints. The first constraint, referred to as a "fixed sample size" constraint, assumes that all the designs initiate with 100,000 live births in the NCS cohort. By constraining the initial sample size in this manner, the implementation costs and the sample size at later stages of the study vary across the 23 different designs under consideration (with costs generally ranging from $2.6B to $3.7B). On the other hand, the second constraint, referred to as a "fixed-cost" design constraint, assumes that all designs must meet an overall study cost of approximately $2.7 billion. By constraining the study resources in this manner, the sample size at both the beginning and at the end of the study will vary across the 23 designs considered. In other words, the number of subjects that can be recruited and followed will depend on the costs associated with each design (e.g., one design may have the financial resources to recruit and follow 90,000 initial participants, whereas another design may only have the financial resources to recruit 70,000 initial participants).

In Chapter 8 of the report we specifically focus on the issue of estimating costs for the study with potential cost differentiators among the four modes of recruitment – National PBS (NPBS), PBS of the geographic area around a Center (area PBS), PBS of Center patients, and purposive sampling of Center patients – resulting in differing cost estimates based on the proportion of the NCS cohort recruited from each of these frames and the number of PSUs selected. Cost estimates were developed within each of seven major activity areas (outlined in Chapter 4) for each of the selected design options, and cost differentiators among the four sampling approaches were identified. (Note that caution should be used in interpreting these cost estimates as many assumptions were made regarding retention rates, number and frequency of samples obtained from participants, and operational and management costs over a 25-year period.) The following general conclusions were apparent in the cost estimates:

- Measurement-related costs represented the largest expense in the cost model.

- Assumed retention rates can play a significant role in the cost estimates due to decreases in the number of participants, or subjects dropping out of the study, resulting in decreases in the corresponding data collection costs.
- Increasing from 50 to 100 PSUs generally increases costs by approximately 10 percent. This indicates the tradeoff between the desire to select individuals in a larger number of locations (i.e., perhaps resulting in a more geographically diverse sample with a broader range of exposures), and the financial costs associated with collecting data in a larger number of geographic areas.

In addition to estimating the costs associated with each of the 23 designs, we also calculate the power of each design to detect relationships of interest. As discussed in Chapter 9 of the report, for a study like the NCS, with multiple hypotheses and multiple inferences of interest, there are many ways to assess power (e.g., different statistical tests, alternative models, different inference goals), and there are many factors that influence the calculation of power (e.g., prevalence of the outcome, strength of the exposure/outcome relationship, etc.). Thus, the power calculations presented in the report focus on a number of relatively simple models relating a categorical exposure variable (exposed/unexposed) to a categorical health outcome (present/absent) and motivated by the core hypotheses of the study. In particular, for each of the 23 designs, a total of nine hypotheses (spanning a range of life stages and alternative disease and exposure occurrence rates) were investigated, and the power to detect the relationship of interest for each of the complex designs under the selected model was calculated via simulation for varying degrees of the strength of the exposure/outcome relationship, for weighted and unweighted analyses[3], for the two different design constraints discussed above (fixed costs versus fixed sample size), and for the two different retention rate assumptions. While the conclusions are often specific to a selected hypothesis or inference goal, the following general conclusions were identified:

- For less common outcomes, for outcomes assessed later in life, and for less common exposures, only stronger exposure/outcome relationships (i.e., only larger odds ratios) are detectable with sufficient power.
- For unweighted (model-based) analyses, it is generally the case that the design that provides the largest available sample size at the selected life-stage corresponds to the design with the highest power. In other words, the design with the largest retention rate corresponds to the design with the highest power.
- Comparing the power for a weighted analysis to that for an unweighted analysis, many of the designs indicate a larger effect of unequal weighting across the cohort (at least larger than the effect of clustering).
- In general, for the weighted analyses, power is enhanced by including a larger fraction of the cohort sampled probabilistically from relatively unrestricted populations. In other words, designs with the largest available sample among the group of people with the largest sampling weights (i.e., the smallest probability of selection) correspond to designs with the highest power.

[3] A weighted analysis is an analysis that incorporates the subject-specific sampling weights and allows inferences to be applied to the wider sampling frame population, whereas an unweighted analysis treats all individuals as equally weighted and allows inferences to be applied to the population of subjects included in the cohort. Using model-based assumptions, results from unweighted analyses may also be generalized to similar populations.

- For the fixed sample size designs, there is relatively little difference in power from the weighted analyses when using a 50 PSU design versus a 100 PSU design (assuming the same proportion of the cohort is selected in the NPBS), suggesting that there is only a small effect of clustering for PSU sizes on the order of 100.
- On the other hand, for the fixed cost designs, the 50 PSU design provides greater power from both unweighted (model-based) and weighted analyses than the 100 PSU design (again, assuming the same proportion of the cohort is selected in the NPBS) due to its lower costs and resulting ability to follow a larger cohort of children. However, the 50 PSU design may pose other feasibility challenges with respect to recruiting a larger number of participants (especially in rural areas) and following participants who move.

As expected, the results of these power calculations indicate that the "optimal" design from the power perspective (i.e., the design with highest power) depends on many factors. For example, the weighted analysis power for hypotheses assessed early in life is generally the highest when the portion of the cohort sampled in the NPBS is largest; however, the unweighted analysis power for hypotheses assessed later in life is generally the highest when the portion of the cohort sampled in the NPBS is smallest. In other words, determination of the "optimal" design will depend on the relative importance of the different NCS objectives (e.g., the relative importance of the hypotheses, the relative importance of the different inference goals, etc.).

CONCLUSIONS

The premise upon which the report is based is that there are multiple legitimate design options for the NCS, each having their own strengths as well as their own limitations in terms of meeting the study objectives. As summarized above, the majority of the report is dedicated to providing estimates for some of the more difficult-to-assess and data-dependent properties of the design options including implementation, costs, recruitment and retention rates, and statistical power. In looking across all this information on the plausible characteristics associated with various design options, and in evaluating that information relative to the study objectives (i.e., applying the criteria for assessing design options described in Chapter 10 and Appendix B1), the first and most important conclusion is that significant tradeoffs appear inevitable. Some of these important tradeoffs include:

- Differences in retention rates associated with individuals selected from different sampling frames (and the corresponding inefficiencies of following individuals that drop out of the study) balanced against the desire for a nationally representative sample of subjects.
- Potential increases in costs associated with recruiting women prior to pregnancy (and following those women over the period of recruitment) versus the potential loss of important pre- and/or peri-conception information.
- Higher potential to satisfy internal validity but less potential to satisfy external validity for designs with higher retention rates and a smaller portion of the cohort selected from the largest sampling frame populations (see Appendix A for a discussion of internal and external validity).

- Higher potential to satisfy external validity with a national probability based sample but less potential to satisfy the need for community involvement and/or the need for specialized measures (e.g., if Centers are not involved in the process).
- National PBS approach provides greater ability to generalize to larger population and perhaps a greater resource for future studies on the basis of protection against bias. On the other hand, the less restrictive recruitment standards in a Center-based approach may foster increased retention rates and allow more information on covariates to help serve as a resource for future studies.

These apparent tradeoffs lead to the general conclusion that there is not a single design that clearly distinguishes itself as the best choice from all perspectives. In light of the study givens, it appears that a final design that includes the involvement of academic medical centers would satisfy the community involvement and specialized measure requirements of the NCS. On the other hand, it appears that including a probability component offers many advantages related to external validity. Therefore, a hybrid approach within the family of designs that incorporates both sampling approaches seems highly desirable. For example, hybrid approaches that include some portion of the sample being conducted as a NPBS, with the remaining percentage covered under centers with the probability component to be negotiated, do offer an attractive balance, achieving power for external validity that appears reasonable for many hypotheses, and that still allow significant community involvement and ability to recruit highly motivated participants. Thus, a hybrid design is possible which is both acceptable and defensible across multiple objectives.

In reviewing the technical information in the report, there remain a number of avenues and open questions that warrant further consideration and investigation in order to better determine the appropriate NCS design. For example, the uncertainty associated with expected recruitment and retention rates associated with different modes of recruitment is one of the most significant limiting factors in more precise estimates of the value of the different designs. This leads to a recommendation for further work (e.g., further examining experiences from other studies to learn more about the factors effecting retention, including additional relevant studies in the estimation of retention rates, etc.) to better understand and/or estimate retention rates for the NCS. Several other important questions/issues that must be considered before making a design decision include:

- Are there alternative sampling frames and/or organizational structures that should be considered and evaluated?
- If the family of designs approach is utilized, what is the optimal allocation of subjects to the NPBS and to the Centers portions of the cohort?
- Is it acceptable to integrate the NPBS into a Centers/Hospital based approach by selecting regions purposively (e.g., by selecting regions that have a capable Center and have desirable characteristics with respect to the goals of the study), and recruiting via the household model within these regions.
- What are the important benefits of community involvement and can it be achieved under a large, widely dispersed NPBS?
- How many geographic regions should be considered, and what are the important advantages and disadvantages associated with selecting (either purposively or

probabilistically) a larger number of regions? Should there be a lower bound on the number of subjects selected in any particular region so that the inefficiency of "covering" a certain region for only a small number of subjects is avoided?

- How should the sample design be stratified (e.g., geographic strata, urbanicity strata, racial and ethnic strata, etc.), and are there important populations that should be oversampled?

Ultimately, the choice of design cannot be reduced to a one-dimensional optimization problem. As such, selection of the NCS design cannot be separated from value judgments related to the importance of the different, and sometimes competing, study objectives (see Appendix A for further discussion of the differing perspectives). The information in the report, however, allows decision makers to understand those tradeoffs in detail, and therefore to be able to make informed decisions when choosing one design over another by understanding what is being gained and what is being lost.